Mejora de los sistemas multimodales mediante el uso de ganancia de información

Manuel Carlos Díaz Galiano
Universidad de Jaén
Campus Las Lagunillas, Edif. A3. E-23071
mcdiaz@ujaen.es

Mª Teresa Martín Valdivia
Universidad de Jaén
Campus Las Lagunillas, Edif. A3. E-23071
maite@ujaen.es

Arturo Montejo Raez
Universidad de Jaén
Campus Las Lagunillas, Edif. A3. E-23071
amontejo@ujaen.es

L. Alfonso Ureña López
Universidad de Jaén
Campus Las Lagunillas, Edif. A3. E-23071
laurena@ujaen.es

Resumen: En este trabajo se discute la utilización de la ganancia de información (IG) para reducir y mejorar la información textual incluida en los sistemas de recuperación de información multimodal. Además se muestran los distintos experimentos realizados combinando esta técnica de reducción con la mezcla de información visual y textual, para comprobar que la información textual consigue mejorar los sistemas multimodales convencionales.
Palabras clave: Recuperación de Información Multimodal, Ganancia de Información, Corpus médicos multimodales

Abstract: This paper discusses the use of information gain (IG) to reduce and improve the textual information included in multi-modal information retrieval systems. Furthermore, a number of experiments are described that combine this reduction technique with a visual- and textual-information merge. These show that the textual information manages to improve conventional multi-modal systems.
Keywords: Multimodal Information Retrieval, Information Gain, Medical Multimodal Corpus

1 Introducción

La ingente cantidad de información disponible electrónicamente en cualquier formato pone de manifiesto la necesidad de desarrollar técnicas que permitan acceder a dicha información de una manera eficiente. Actualmente, la información disponible electrónicamente tiende a ser cada vez más multimodal, incluyendo cualquier tipo de información. La adición de imagen y sonido a los sistemas informáticos suponen un gran avance tecnológico desde el punto de vista del usuario puesto que la comunicación humana es intrínsecamente multimodal (incluye sonidos, textos, fotografías, imágenes en movimiento…) (Lewis et al, 2006). Sin embargo, sería un error pensar que simplemente el tener más información, aunque esta información sea multimodal, puede resolver los problemas de acceso a la misma de manera eficiente. Todo lo contrario, si no disponemos de sistemas que sean capaces de realizar una recuperación eficaz, no importará la calidad de la información disponible puesto que no seremos capaces de acceder a ella aunque esté ahí.

Los sistemas de recuperación de información visual o sistemas de recuperación de imágenes basados en contenido, han sido denominados de diversas formas: sistemas CBIR (Content Based Information Retrieval), CBVIR (Content Based Visual Information Retrieval) o QBIC[1] (Query by imagen content), este último fue el nombre que IBM dio a su primer sistema implementado en los años 90. Un sistema CBIR es una aplicación que busca dentro de una colección de imágenes aquellas que son semejantes o que tienen un contenido similar a una imagen dada como consulta. Que

[1] http://wwwqbic.almaden.ibm.com/

 © Sociedad Española para el Procesamiento del Lenguaje Natural

dichos sistemas sean *basados en contenido* significa que la búsqueda se realiza basándose en las características y el contenido de la imagen y no en otro tipo de información añadida manualmente, como por ejemplo el título de la imagen o palabras clave[2]. La primera vez que se utilizó el término CBIR fue por Kato (1992), para describir sus experimentos donde realizaba una recuperación visual basándose en los colores y las formas de las imágenes.

Actualmente, están generando bastante interés sistemas en los que además de almacenar imágenes se incluye cierto texto asociado a dichas imágenes (meta-datos). Es el caso, por ejemplo, de los expedientes médicos en los que una radiografía puede tener asociada una información textual relativa al historial clínico del paciente, al comentario de un especialista sobre la radiografía, información sobre el tratamiento propuesto al paciente… Otro ejemplo sería una colección de fotografías con comentarios sobre las mismas. Las fotografías pueden ser cuadros de un museo, fotografías asociadas a noticias en un periódico o catálogos de productos de cualquier tipo. Una manera de recuperar información en este tipo de sistemas podría incluir la recuperación visual por una parte, la recuperación textual por otra, y finalmente, una mezcla de resultados parciales (visuales y textuales) que persigan la optimización de la respuesta dada.

Un ejemplo práctico de la utilización de un sistema mixto (CBIR+IR), lo tenemos en el trabajo diario de un médico. Éste posee casos clínicos de sus pacientes. Dichos casos están compuestos por textos descriptivos del caso e imágenes que ilustran la dolencia. Con un sistema CBIR, ayudado por un sistema IR, como el que se muestra en la Figura 1, el médico podría utilizar una imagen de una dolencia (por ejemplo, una radiografía) y obtener información de casos similares a dicha dolencia. Por lo tanto, la recuperación sería tanto visual como textual, ya que los casos están compuestos tanto por información textual del caso como por imágenes.

Cabe pues plantearse que una recuperación eficiente del texto puede ayudar a mejorar la calidad de los sistemas multimodales en general. El texto puede beneficiarse de las imágenes y viceversa. De hecho, así se pone de manifiesto en distintos foros y conferencias realizadas en los últimos años (Clough et al., 2006, Declerck et al., 2004, Müller et al., 2006).

En una colección con gran cantidad de metadatos nos encontramos con la problemática de elegir aquellos metadatos que son de mayor utilidad y desechar aquellos que pueden añadir información no relevante (ruido) en nuestro

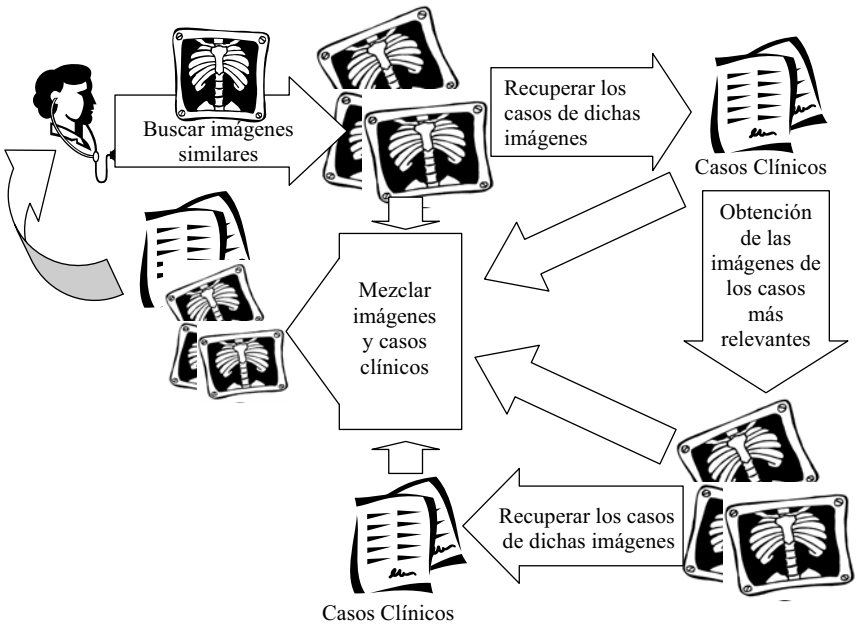

Figura 1: Ejemplo de utilización de un sistema mixto CBIR + IR

[2] http://en.wikipedia.org/wiki/CBIR

sistema. La ganancia de información es una técnica a través de la cual podemos seleccionar aquellos metadatos que aportan mayor información al sistema ignorando aquellos que no sólo no aportan información alguna, sino que en ocasiones incluso introducen ruido y pueden distorsionar la respuesta del sistema.

En este trabajo se propone el uso de la ganancia de información como técnica para mejorar la calidad del corpus textual asociado a una colección de documentos que representan expedientes médicos. Una mejora del corpus textual implica una mayor eficacia en la recuperación de este tipo de información, lo que repercute directamente en la eficacia del sistema multimodal global.

El resto de este artículo se organiza de la siguiente manera. En primer lugar, se hace una breve introducción a la ganancia de información, indicando su formulación y sus principales aplicaciones. A continuación se describe la colección de documentos que se ha utilizado para realizar los experimentos. En el apartado 4 se explica cómo se ha utilizado la ganancia de información para seleccionar las etiquetas con mayor información sobre la colección de documentos multimodales. En el apartado 5, se presentan los experimentos realizados sobre la colección multimodal así como los resultados obtenidos. Por último, se muestran los resultados obtenidos y se presentan las conclusiones junto con un avance sobre la orientación de los trabajos futuros que dan continuidad a esta investigación.

2 Ganancia de Información

La Ganancia de Información (Information Gain – IG) es una medida basada en la entropía de un sistema, es decir, en el grado de desorden de un sistema (Shannon, 1948). Esta medida nos indica cuánto se reduce la entropía de todo el sistema si conocemos el valor de un atributo determinado. De esta forma, podemos conocer cómo se relaciona el sistema completo con respecto a un atributo, o lo que es lo mismo, cuánta información aporta dicho atributo al sistema.

La fórmula para calcular la IG es la siguiente:

$$IG(C|E) = H(C) - H(C|E) \quad (1)$$

donde

- IG(C|E): es la ganancia de información de la etiqueta o característica E,
- H(C): es la entropía del sistema
- H(C|E): es la entropía relativa de sistema conocido el valor de la etiqueta E.

La entropía del sistema nos indica el grado de desorden del mismo y viene dada por la siguiente fórmula:

$$H(C) = -\sum_{i=1}^{|C|} p(c_i)\log_2 p(c_i) \quad (2)$$

donde $p(c_i)$ es la probabilidad del valor *i*.

La entropía relativa se calcula de la siguiente manera:

$$H(C \mid E) = \sum_{j=1}^{|E|} p(e_j)\left(-\sum_{i=1}^{|C|} p(c_i \mid e_j)\log_2 p(c_i \mid e_j) \right) \quad (3)$$

donde $p(e_i)$ es la probabilidad del valor *i* para la característica *e*, y $p(c_i|e_j)$ es la probabilidad de c_i relativa a e_j

La principal aplicación de la IG es la selección de características. Por lo tanto, es un buen candidato para la selección de aquellos meta-datos que son útiles para el dominio en el que se usa la colección.

La IG se ha empleado en multitud de estudios (Quinlan, 1986), la mayoría de ellos de clasificación. Algunos ejemplos son la categorización de textos (Text Categorization – TC) (Yang y Pedersen, 1997), aprendizaje automático (Machine Learning – ML) (Mitchell, 1996) o detección de anomalías (Anomaly Detection – AD) (Lee y Xiang, 2001).

Nosotros partimos de una colección multimodal que representa informes médicos consistentes en a un conjunto de imágenes médicas, y a cada una de ellas se asocia información textual mediante diferentes etiquetas (meta-datos) algunas de las cuales no aportan apenas información. Por ejemplo, este es el caso de la etiqueta LANGUAGE, ya que esta etiqueta contiene el mismo valor para toda la colección. Con la finalidad de depurar y mejorar la calidad del corpus textual, hemos calculado la ganancia de información de las etiquetas para poder realizar una selección de aquellas que aporten una información más discriminante.

3 Descripción de la colección multimodal

Para realizar los experimentos se ha utilizado la colección suministrada por la organización de la

competición CLEF (Cross Language Evaluation Forum)[3] en la tarea concreta sobre recuperación de imágenes médicas (Müller et al., 2006). Esta tarea se conoce como ImageCLEFmed[4]. La colección de documentos proporcionada para esta subtarea está formada por 4 subcolecciones de datos: CASImage, Pathopic, Peir y MIR, e incluyen unas 50,000 images.

Cada subcolección se organiza en "casos" (véase Figura 2). Un caso está formado por una o varias imágenes (dependiendo de la colección) y un conjunto de anotaciones en formato texto asociadas a dicha imagen. Las anotaciones están marcadas con etiquetas y constituyen los metadatos de la colección. Algunos casos incluyen también otras imágenes relacionadas con el caso. Por ejemplo, se puede tener una imagen de una radiografía de un femur, y asociada a esta imagen disponer de otras que muestren secciones del mismo femur, una resonancia magnética, una fotografía, etc.

La colección CASImage[5] contiene unas 8.725 imágenes agrupadas en 2.076 casos. Esta colección está compuesta de imágenes de escáner, rayos x, ilustraciones, fotografías y presentaciones. El 20% de los casos está en inglés y el resto en francés. La colección MIR (Mallinckrodt Institute of Radiology)[6] contiene 1.177 imágenes de medicina nuclear repartidas en 407 casos. Cada caso contiene anotaciones en inglés. Los casos de la colección PEIR (Pathology Education Instructional Resource)[7] sólo contienen una imagen por caso. Dicha colección contiene 32.319 imágenes con sus respectivos casos anotados en inglés. La información sobre las imágenes es muy escasa, aunque está bien clasificada en campos. La colección PathoPIC[8] contiene 7.805 imágenes de patologías. Al igual que la colección PEIR, existe una sola imagen por caso, aunque cada caso está anotado en dos idiomas, alemán e inglés. El idioma original de los casos es el alemán, por lo que las anotaciones en inglés son traducciones de dichos casos.

Para generar la colección textual se utiliza un fichero índice que permite determinar qué imágenes y anotaciones textuales pertenecen a cada caso[9]. Las anotaciones textuales están en formato XML y la mayoría se encuentran escritas en inglés, sin embargo, el 80% de la subcolección CASImage está etiquetada en francés. Esto implica que antes de preprocesar

Metadatos del caso

```
<ID>3349</ID>
<Description>On the frontal and lateral chest x-rays, perivascular haziness
is visible with a ground glass and diffuse nodular
infiltrate.</Description>

<Diagnosis>Acute eosinophilic pneumonia</Diagnosis>

<ClinicalPresentation>Patient with a fever and respiratory insufficiency
since 5 days.</ClinicalPresentation>

<Commentary>The diagnosis was based on a bronchoscopy with bronchoalveolar
lavage, demonstrating eosinophilia > 25%, as well as the absence of
parasites or any other pathogen.

…
```

Imágenes

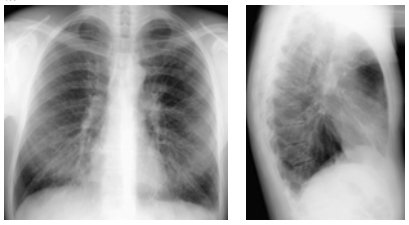

Figura 2: Ejemplo parcial de un caso de la colección CASImage

[3] http://www.clef-campaign.org/
[4] http://ir.ohsu.edu/image/
[5] http://www.casimage.com
[6] http://gamma.wustl.edu/home.html
[7] http://peir.path.uab.edu
[8] http://alf3.urz.unibas.ch/pathopic/intro.htm
[9] Para más información de la organización de la colección consultar la página del CLEF (http://ir.ohsu.edu/image/2005protocol.html) .

la colección completa es necesario realizar la traducción automática de las anotaciones del francés al inglés. Para ello se ha utilizado un traductor automático a través de Internet. Concretamente, se ha utilizado el traductor on-line Reverso[10]. La colección Pathopic contiene anotaciones en inglés y en alemán pero el corpus es paralelo (las mismas anotaciones en inglés están también en alemán). En este caso, simplemente se han ignorado las anotaciones en alemán y sólo se han incorporado a la colección completa las anotaciones en inglés. Algunos casos (aunque muy pocos) no contienen ninguna anotación. La calidad de los textos de las colecciones varía de una subcolección a otra, e incluso dentro de la misma subcolección.

Se considera que por cada imagen se tiene un documento textual con las anotaciones sobre el caso. Si un caso tiene más de una imagen asociada, el texto del caso se repite tantas veces como imágenes contenga, tal y como se muestra en la Figura 3. De esta manera, se genera la colección textual completa con todos los documentos de cada una de las subcolecciones.

Tomando como ejemplo de partida el caso de la Figura 2, la descomposición se realizaría de acuerdo al esquema mostrado en la Figura 3.

4 Selección de etiquetas

Para depurar y mejorar la calidad de la colección de documentos, se ha aplicado la ganancia de información con el fin de permitir la selección de las mejores etiquetas y eliminar aquellas que no aportan apenas información. Para ello, se ha calculado la IG para cada una de las etiquetas de cada subcolección. Puesto que cada subcolección

CASImage, Pathopic, Peir y MIR tiene un conjunto de etiquetas diferente, la IG se calcula en el ámbito de cada subcolección, independientemente del resto. Si tomamos la formula (1), C sería el conjunto de casos y E el conjunto de posibles valores de la etiqueta XML de nombre E.

Para calcular el valor de IG, se calcula la entropía del conjunto de casos C como:

$$H(C) = -\sum_{i=1}^{|C|} p(c_i)\log_2 p(c_i) = \\ = -\sum_{i=1}^{|C|} \frac{1}{|C|}\log_2 \frac{1}{|C|} = -\log_2 \frac{1}{|C|} \quad (4)$$

Y la entropía del conjunto de casos C condicionada por la etiqueta E como:

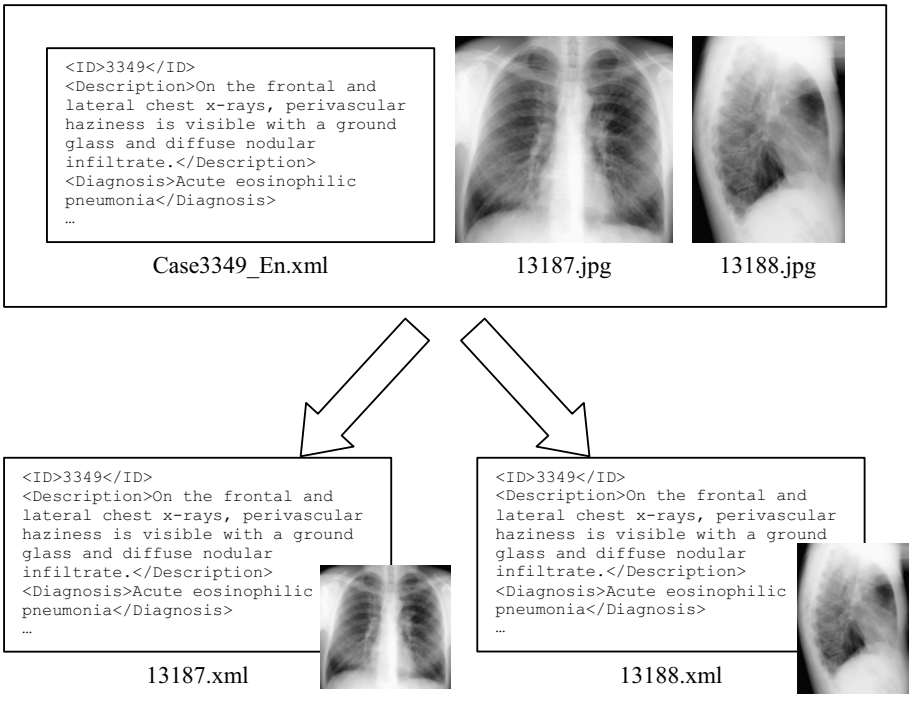

Figura 3: Extracción de la anotación textual de cada imagen

[10] http://www.reverso.net

$$H(C \mid E) = \sum_{j=1}^{|E|} \frac{|C_{e_j}|}{|C|} \left(-\sum_{i=1}^{|C|} \frac{1}{|C_{e_j}|} \log_2 \frac{1}{|C_{e_j}|} \right) = \quad (5)$$

$$= -\sum_{i=1}^{|C_{e_j}|} \frac{|C_{e_j}|}{|C|} \log_2 \frac{1}{|C_{e_j}|}$$

donde C_{ej} es el subconjunto de casos en C que tienen el valor e_j en la etiqueta E. El valor de e_j es el conjunto de palabras que forman dicha etiqueta (sin considerar el orden de las palabras). Conociendo la entropía del sistema y la entropía condicionada podemos componer la ecuación final de la siguiente manera:

$$IG(C \mid E) = -\log_2 \frac{1}{|C|} + \sum_{i=1}^{|C_{e_j}|} \frac{|C_{e_j}|}{|C|} \log_2 \frac{1}{|C_{e_j}|} \quad (6)$$

Se calcula la ganancia de información para cada una de las etiquetas en cada una de las colecciones. Una vez que cada etiqueta tiene asociada su IG, se ordenan utilizando este valor como referencia. A continuación, la colección final se crea seleccionando aquellas etiquetas que tienen los valores de IG más altos. No obstante, existen etiquetas dentro de la colección (como por ejemplo el campo identificador ID) con un valor de IG muy alto y cuyo contenido es poco representativo, ya que difiere enormemente para cada caso y el número de términos que contiene es muy pequeño. Por lo tanto, antes de ordenar y seleccionar las mejores etiquetas, se eliminan aquellas cuya frecuencia media de palabras en la subcolección sea inferior a un umbral. De esta forma, una colección generada utilizando el contenido del 100% de la etiquetas con mejor IG contendrá, no obstante, menos etiquetas (y por lo tanto, menos texto) que una colección con todas la etiquetas.

5 Experimentos y resultados

El objetivo principal que se persigue es demostrar que los resultados obtenidos con un corpus en el que se han filtrado aquellas etiquetas que aportan poca información (es decir, con una IG baja) son mejores que cuando se utiliza el corpus completo. Para ello se han realizado experimentos utilizando diferente número de etiquetas seleccionadas. Concretamente, se han tomado etiquetas con la mayor IG de 10 en 10 por ciento sobre el total, empezando en el 10% hasta el 100% de las etiquetas. También se han realizado experimentos con una colección que utiliza todas las etiquetas (sin aplicar el filtro por frecuencias comentado anteriormente).

Además de la colección multimodal, la organización del CLEF también pone a disposición de los participantes 25 consultas compuestas por una o varias imágenes y por un texto asociado.

5.1 Casos base visual y textual

Para poder analizar las mejoras que el sistema híbrido propuesto pudiera aportar, se han realizado dos casos experimentales que sirven de base: un caso basado únicamente en las imágenes, y otro en la información textual.

Como caso base visual se ha tomado el resultado obtenido para cada consulta utilizando exclusivamente un sistema CBIR (es decir, sin tener en cuenta el texto sino únicamente haciendo uso de la imagen). Para ello, se han utilizado las listas de resultados suministrada por la organización del CLEF para cada una de las 25 consultas. Estas listas (una por consulta) se obtienen como resultado al presentar una imagen a un sistema de recuperación de imágenes denominado GIFT[11] (GNU Image Finding Tool). Se trata de un sistema CBIR que usa 4 características de imagen para realizar la recuperación (Squire et al., 2000). El resultado obtenido tras una consulta con una imagen al sistema GIFT consiste en una lista de imágenes ordenadas según su valor de relevancia con respecto a la imagen de consulta.

Como caso base textual se considera el resultado obtenido por cada consulta utilizando el texto de la misma sobre un sistema de recuperación de información textual. El sistema utilizado es LEMUR[12]. Este es un sistema multiplataforma desarrollado como parte del Proyecto LEMUR, una colaboración entre los departamentos de Informática de las universidades de Massachussets y Carnegie Mellon. Dicha herramienta permite el filtrado y la indexación de grandes colecciones documentales y la recuperación de información en dichas colecciones, utilizando una gran variedad de modelos de recuperación. El resultado obtenido tras una consulta a LEMUR con el texto de cada una de las 25 consultas es

[11] http://www.gnu.org/software/gift/
[12] http://www.lemurproject.org/

una lista de documentos ordenados por su valor de relevancia.

5.2 Expansión de las consultas textuales

Para mejorar los resultados de los casos base se ha utilizado la información textual disponible de cada caso y aplicando un método de retroalimentación. De este modo, hemos expandido las consultas originales con el texto asociado a las 4 primeras imágenes recuperadas con el sistema GIFT. El texto utilizado para realizar la expansión depende de la colección donde se realiza la recuperación de información textual (10%, …, 100% o todas).

5.3 Mezcla de resultados textuales y visuales

Además de los casos base textual y visual, se han realizado 3 tipos de experimentos:

- **Solo texto y GIFT:** La forma más sencilla de incorporar información visual al resultado final consiste en mezclar el caso base textual con el caso base visual dando distintos pesos a los valores de relevancia (RSV) de ambos casos (Figura 4). La fórmula sería la siguiente:

$$RSV_{final} = (RSV_{text} \cdot \alpha) + (RSV_{visual} \cdot \beta) \quad (5)$$

  donde α y β son los pesos de cada lista y cumplen que $\alpha+\beta = 1$
- **Consulta textual expandida:** Otra manera de mezclar los resultados textuales y visuales es utilizando la lista obtenida al expandir la consulta textual. De esta forma, la aportación visual al experimento es mayor (Figura 5).
- **Consulta textual expandida y GIFT:** Por último, se puede mezclar la lista de la consulta expandida con la lista del GIFT, utilizando la fórmula (5), para realizar una doble aportación visual.

5.4 Resumen de experimentos

Cada uno de los experimentos diseñados () se ha lanzado contra cada una de las colecciones generadas usando filtrado de etiquetas con IG. A dichas colecciones se le ha denominado según el porcentaje de etiquetas seleccionadas: *Coll_10, Coll_20, ..., Coll_100*. Al corpus completo con todas las etiquetas se le ha denominado *Coll_All*. Recordemos que los corpus con el 100% de las etiquetas y con todas las etiquetas no son iguales.

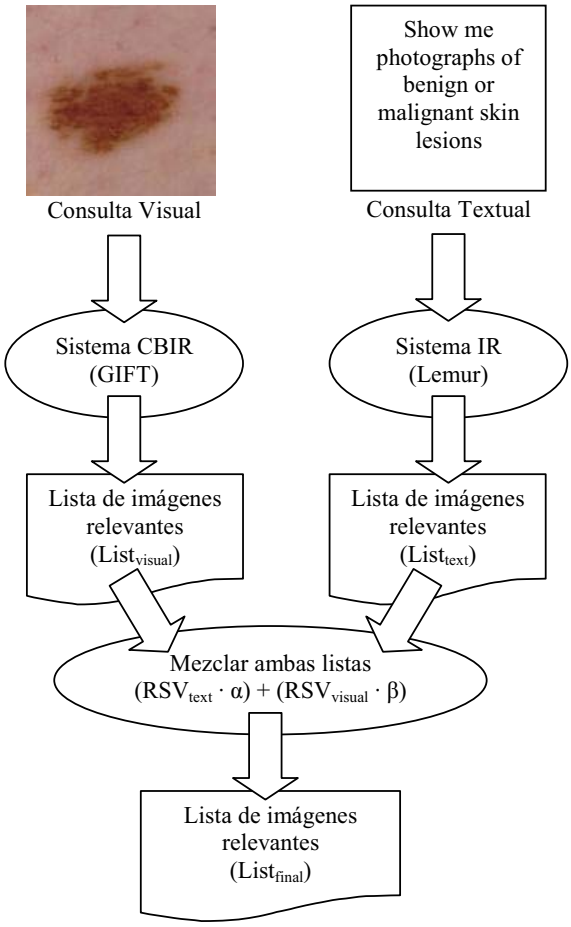

Figura 4: Esquema de mezclado de las listas visuales y textuales

Una vez calculada la IG de cada etiqueta y antes de ordenarlas por IG para seleccionar las etiquetas que tienen mayor valor, se han eliminado aquellas etiquetas cuya frecuencia media de palabras es inferior a un umbral dado. De esta forma, eliminamos aquellas etiquetas que teniendo pocas palabras (es decir, poca información) tienen un valor de IG alto. Así pues, el corpus denominado Coll_100 filtrado por IG contiene el 100% de las etiquetas que han superado el umbral de corte, y por lo tanto dicho corpus contiene menos etiquetas que el corpus completo (Coll_All).

Para dar nombre a los experimentos de mezcla de listas se ha optado por la siguiente nomenclatura:

T*alfa*_C*coleción*

(para los experimentos de mezcla)

donde:

- *alfa*: el porcentaje dado al RSV textual
- *colección*: porcentaje de etiquetas que tiene la colección donde se realiza la recuperación textual

Por ejemplo, si un experimento se nombra T90C30, significa que se le ha dado un 90% de importancia al RSV textual (y en consecuencia un 10% al RSV visual) y que se ha utilizado la colección con el 30% de etiquetas con mejor IG.

Para los experimentos donde se realiza expansión de la consulta con las 4 primeras imágenes del GIFT, los experimentos se han nombrado de la siguiente manera:

Expand_C*coleción*
(para los experimentos de expansión)

donde *colección* es el porcentaje de etiquetas que tiene la colección donde se realiza la recuperación de información. Por ejemplo, un experimento llamado ExpandCall, significa que se ha utilizado la colección con todas las etiquetas (all) para realizar la recuperación de información.

En cuanto a los experimentos donde se realiza expansión de la consulta con las 4 primeras imágenes del GIFT más la mezcla de dichos resultados con el caso base textual, los experimentos se han nombrado de la siguiente manera:

ExpandT*alfa*_C*coleción*
(para los experimentos de expansión)

Por ejemplo, un experimento con nombre ExpandT50C20, significa que se le ha dado un 50% de importancia al RSV textual expandido (y en consecuencia un 50% al RSV visual) y que se ha utilizado la colección con el 20% de etiquetas con mejor IG.

Experimento	α (porcentaje textual)	Colección utilizada
GIFT (caso base visual)	0%	Ninguna textual
OnlyText (caso base textual)	100%	10%, …, 100%, all
Texto expandido con GIFT	10%, …, 100%	10%, …, 100%, all

Tabla 1: Resumen de experimentos realizados.

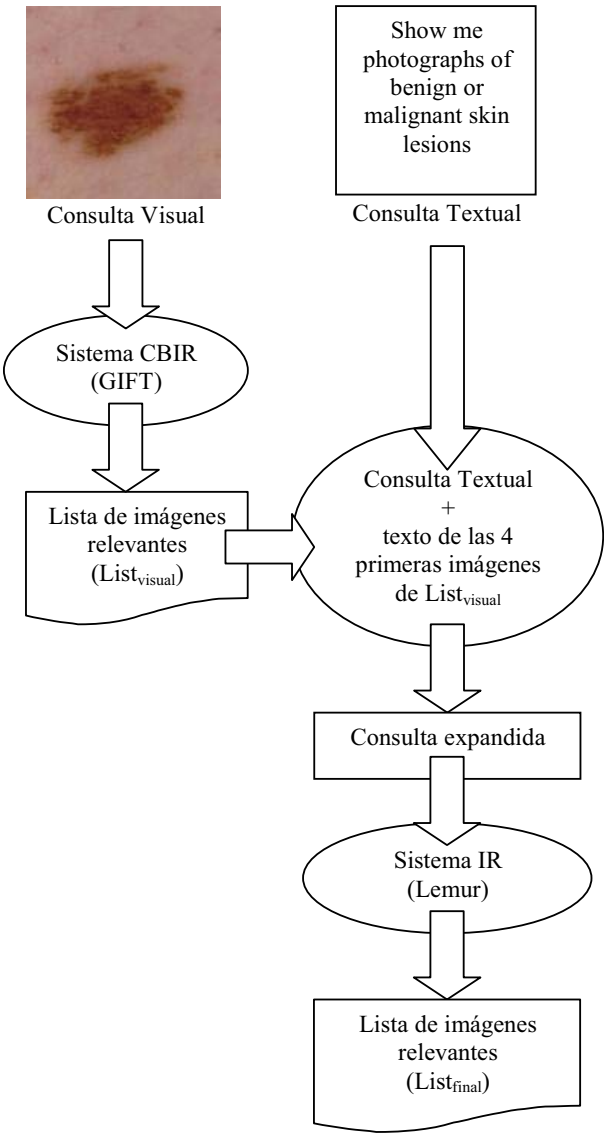

Figura 5: Esquema de expansión de la consulta utilizando las 4 primeras imágenes obtenidas con GIFT

5.5 Resultados

Dependiendo del experimento, tenemos varios tipos de resultados

5.5.1 Sólo texto con diferentes colecciones

Con los primeros resultados obtenidos, podemos comparar cómo se comportan las distintas colecciones generadas, es decir, aquellas colecciones que tienen diferente porcentaje de etiquetas (etiquetas elegidas según su IG).

Como se puede comprobar en la Figura 6, al utilizar sólo las consultas textuales para recuperar las imágenes relevantes, se obtienen mejores resultado que utilizando únicamente el

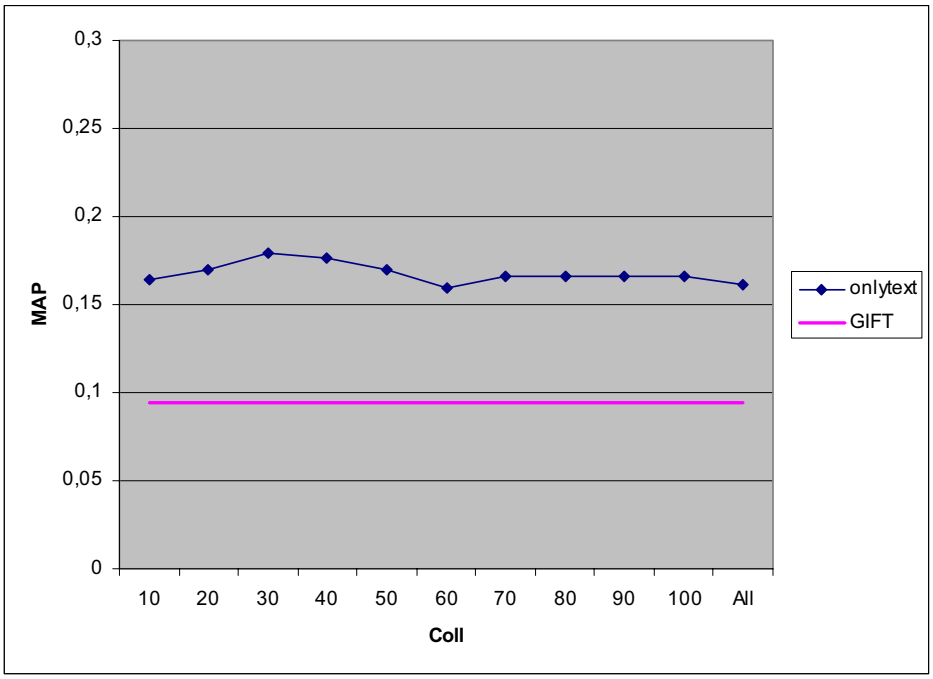

Figura 6: Comparación entre los casos base visual y textual

sistema GIFT[13], obteniéndose una precisión media (MAP) de casi el doble en el mejor de los casos (usando la colección con el 30% de etiquetas).

En términos generales, las colecciones que tienen un porcentaje de etiquetas reducido (entre el 20% y el 50%) obtienen los mejores resultados, con un valor de MAP entre 0,18 y 0,17.

5.5.2 Mezcla de solo texto y GIFT

En cuanto a los experimentos realizados mezclando ambas listas (visual y textual), podemos comprobar que aquellos que dan más peso al texto obtienen mejores resultados, tal y como era de esperar, ya que la recuperación textual genera mejores resultados que la recuperación visual con GIFT. No obstante, aquellos experimentos donde el peso dado al texto está entre el 40% y el 90% también consiguen superar al caso base textual (Figura 6).

Los experimentos con mejores resultados son aquellos en los que el peso de la parte textual no es muy elevado (50%, 60% y 70%) lo que efectivamente demuestra que la combinación de los dos tipos de resultados (textual y visual) permite superar los resultados obtenidos de manera independiente (Tabla 2).

El uso de colecciones con un menor número de etiquetas también mejora los resultados obtenidos (Figura 7). En este caso, podemos comprobar cómo se acentúa la influencia que produce la cantidad de etiquetas de la colección. El experimento que mejor se comporta es aquel que da un 60% de importancia al texto (α=0,6; β=0,4). En este experimento se comprueba que utilizar una colección que posea un número de etiquetas reducido (entre el 20% y el 40%) mejora la calidad de las soluciones.

Como se puede comprobar, la mezcla de resultados supera con creces los resultados visuales (GIFT), incluso en aquellas mezclas en los que los resultados están por debajo del caso base textual.

5.5.3 Consulta textual expandida

Cuando se genera una nueva consulta con el texto original de la consulta más el texto asociado a las cuatro primeras imágenes de la lista visual, el resultado no difiere mucho de utilizar sólo texto. De hecho, los resultados obtenidos no son nada significativos puesto que prácticamente son iguales a los obtenidos con el caso base textual (la mayor diferencia entre resultados es de 0.001). Por este motivo, no se muestran dichos resultados.

[13] El valor MAP para el GIFT es 0.094

	C10	C20	C30	C40	C50	C60	C70	C80	C90	C100	CAll
onlytext	0,1645	0,1695	0,1791	0,1762	0,1695	0,1599	0,166	0,1659	0,1659	0,1659	0,1614
T10	0,1132	0,1150	0,1161	0,1153	0,1147	0,1144	0,1155	0,1153	0,1154	0,1154	0,1166
T20	0,1309	0,1341	0,1360	0,1347	0,1335	0,1326	0,1342	0,1339	0,1342	0,1342	0,1360
T30	0,1544	0,1581	0,1610	0,1584	0,1568	0,1538	0,1554	0,1552	0,1553	0,1553	0,1572
T40	0,1875	0,1912	0,1965	0,1898	0,1862	0,1792	0,1795	0,1791	0,1791	0,1791	0,1780
T50	0,2073	**0,2115**	**0,2198**	**0,2151**	0,2012	0,1930	0,1970	0,1962	0,1963	0,1963	0,1919
T60	0,2140	**0,2164**	**0,2252**	**0,2238**	0,2074	0,1955	0,2010	0,2001	0,2000	0,2000	0,1995
T70	0,2055	**0,2040**	**0,2131**	**0,2120**	0,1988	0,1885	0,1941	0,1930	0,1933	0,1933	0,1901
T80	0,1922	0,1915	0,2013	0,1995	0,1882	0,1785	0,1851	0,1843	0,1846	0,1846	0,1800
T90	0,1804	0,1825	0,1920	0,1891	0,1806	0,1705	0,1776	0,1768	0,1772	0,1772	0,1716

Tabla 2. Mezcla de solo texto y GIFT

	C10	C20	C30	C40	C50	C60	C70	C80	C90	C100	CAll
onlytext	0,1645	0,1695	0,1791	0,1762	0,1695	0,1599	0,166	0,1659	0,1659	0,1659	0,1614
ExpandT10	0,1131	0,1150	0,1161	0,1151	0,1147	0,1143	0,1155	0,1156	0,1154	0,1154	0,1168
ExpandT20	0,1309	0,1342	0,1362	0,1346	0,1335	0,1327	0,1340	0,1344	0,1341	0,1341	0,1363
ExpandT30	0,1545	0,1582	0,1611	0,1579	0,1566	0,1537	0,1552	0,1554	0,1554	0,1554	0,1581
ExpandT40	0,1876	0,1915	0,1966	0,1894	0,1863	0,1794	0,1793	0,1797	0,1795	0,1795	0,1784
ExpandT50	0,2072	**0,2119**	**0,2202**	**0,2145**	0,2014	0,1931	0,1970	0,1967	0,1964	0,1964	0,1932
ExpandT60	0,2139	**0,2164**	**0,2256**	**0,2228**	0,2073	0,1958	0,2009	0,2006	0,2004	0,2004	0,1997
ExpandT70	0,2063	**0,2043**	**0,2129**	**0,2110**	0,1989	0,1884	0,1935	0,1936	0,1934	0,1934	0,1905
ExpandT80	0,1925	0,1918	0,2013	0,1988	0,1886	0,1785	0,1848	0,1845	0,1847	0,1847	0,1807
ExpandT90	0,1808	0,1828	0,1920	0,1882	0,1806	0,1705	0,1774	0,1774	0,1773	0,1773	0,1729

Tabla 3. Mezcla de la consulta textual expandida y GIFT

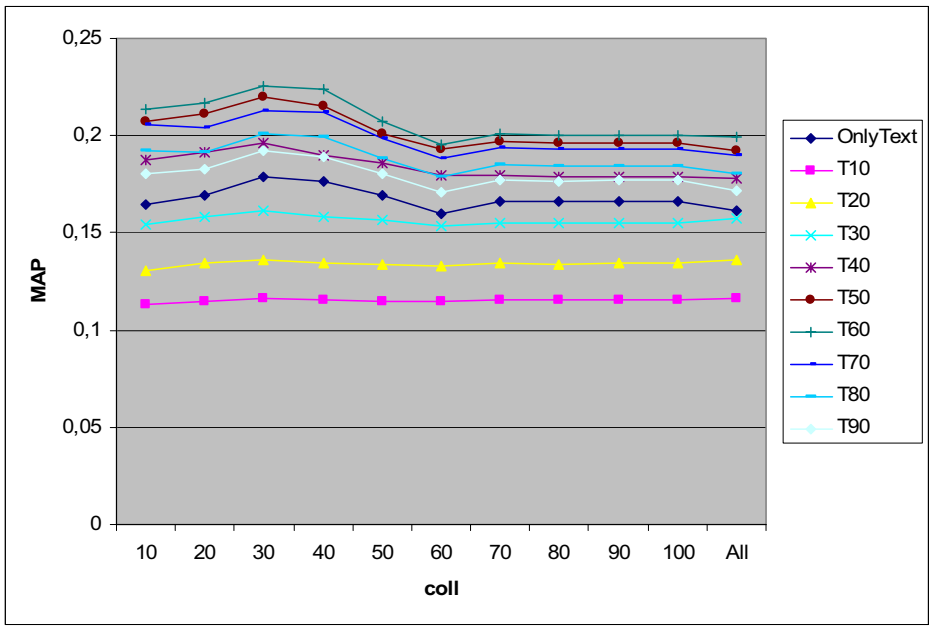

Figura 7: Gráfica comparativa de los distintos métodos de mezclado de listas (visual y textual)

5.5.4 Mezcla de la consulta textual expandida y GIFT

Cuando utilizamos conjuntamente la expansión y la mezcla de listas los resultados son similares a la simple mezcla de listas, ya que como se ha comentado en el apartado anterior, la expansión no mejora los resultados textuales. Sin embargo, el mejor de los resultados global (experimento ExpandT60C30) es levemente superior al mejor de los resultados de la simple mezcla (experimento T60C30), tal y como se puede observar en la Tabla 3. En realidad, si se compara toda la tabla en general, la diferencia es ínfima.

En la Figura 8, podemos observar claramente cómo los mejores resultados se concentran cuando se utilizan colecciones con un porcentaje de etiquetas menor y el peso de la lista textual es superior a la de la lista visual.

6 Conclusiones y trabajos futuros

La selección de etiquetas utilizando el método de IG permite filtrar un corpus con el fin de mejorar la calidad y obtener así mejores resultados en la recuperación de información. Además de reducir el tamaño de los corpus utilizados, este método permite seleccionar aquellas etiquetas más significativas dentro del corpus, o por lo menos, aquellas que más información aportan.

Este sistema de selección no necesita ningún tipo de entrenamiento ni conocimiento externo, simplemente estudia la importancia de cada etiqueta con respecto al total de documentos. Además, es independiente del corpus analizado, ya que en nuestros experimentos el cálculo de la IG se ha realizado de forma independiente en cada subcolección.

Además, se ha comprobado que el uso y combinación de varias fuentes de información (textual y visual) mejora significativamente la utilización de una única fuente. Aunque por una parte, la recuperación textual por si sola supera a la recuperación visual, cuando se utilizan conjuntamente, los resultados superan a los obtenidos con las resuperaciones independientes.

Por último, también se ha comprobado que la expansión de la consulta textual incorporando texto a partir de las imágenes de la recuperación textual no aporta apenas beneficios.

En el futuro se intentará estudiar la incidencia de aplicar esta técnica en sistemas que necesitan más información, como por ejemplo, sistemas de búsqueda de respuestas. Además, se aplicarán todos los resultados obtenidos sobre otras colecciones con metadatos como por ejemplo a las colecciones TRECVid.

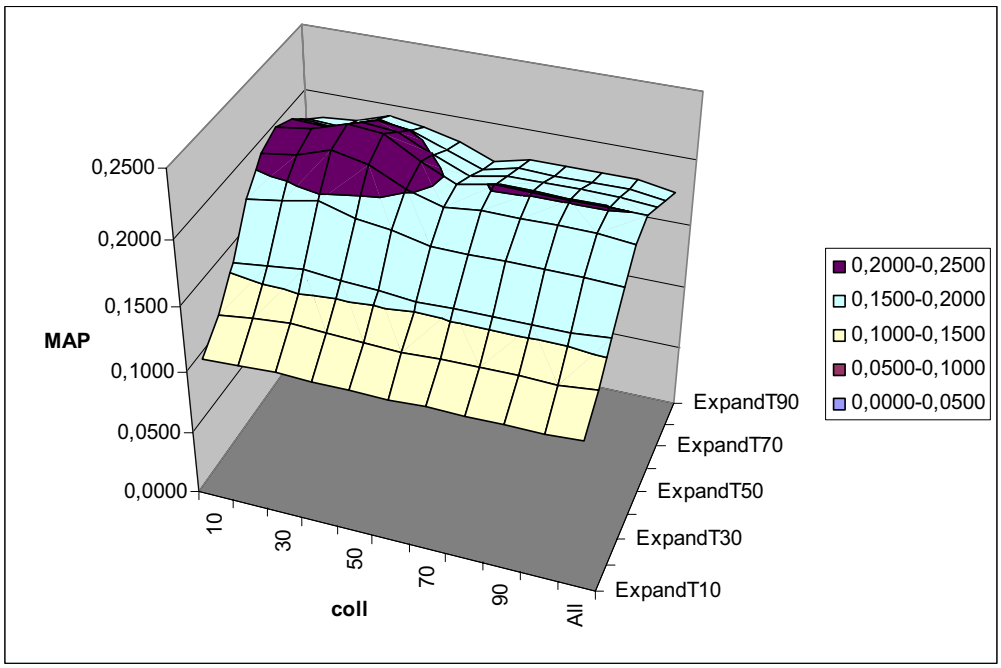

Figura 8: Gráfica comparativa entre los experimentos que utilizan expansión y mezcla de listas

7 Agradecimientos

Queremos expresar nuestra gratitud a la organización del CLEF y más concretamente a Carol Peters, por permitirnos utilizar y referenciar los distintos recursos disponibles en dicho foro.

Este trabajo ha sido parcialmente financiado por el Ministerio de Ciencia y Tecnología a través del proyecto TIMOM (TIN2006-15265-C06-03).

Bibliografía

Clough, P., H. Müller, T. Deselaers , M. Grubinger, T. Lehmann, J. Jensen, W. Hersh. 2005. *The CLEF 2005 Cross-Language Image Retrieval Track*. In Proceedings of the Cross Language Evaluation Forum (CLEF 2005).

Clough, P., M. Grubinger, T. Deselaers, A. Hanbury y H. Müller. 2006. *Overview of the ImageCLEF 2006 photographic retrieval and object annotation tasks*. Evaluation of Multilingual and Multi-modal Information Retrieval – Seventh Workshop of the Cross-Language Evaluation Forum, CLEF 2006.

Declerck, T., J. Kuper, H. Saggion, A. Samiotou, P. Wittenburg y J. Contreras. 2004. *Contribution of NLP to the Content Indexing of Multimedia Documents*. Image and Video Retrieval. LNCS 2004. Volume 3115/2004.

Kato, T. 1992. *Database architecture for content-based image retrieval*. Image Storage and Retrieval Systems, Proc. SPIE 3312, 162-173.

Lee, W., D. Xiang. 2001. *Information-Theoretic Measures for Anomaly Detection*. Proc. of the 2001 IEEE Symposium on Security and Privacy.

Lewis, M. S., N. Sebe, C. Djeraba y R. Jain. 2006. *Content-Based Multimedia Information Retrieval: State of the Art and Challenges*. ACM Transactions on Multimedia Computing, Communications, and Applications, Volume 2. February 2006.

Mitchell, T. 1996. *Machine Learning*. McGraw Hill.

Müller, H., T. Deselaers, T. Lehmann, P. Clough y W. Hersh. 2006. *Overview of the ImageCLEFmed 2006 medical retrieval and annotation tasks*. Evaluation of Multilingual and Multi-modal Information Retrieval – Seventh Workshop of the Cross-Language Evaluation Forum, CLEF 2006. LNCS 2006.

Quinlan, J. R. 1986. *Induction of Decision Trees*. Machine Learning, (1), 81-106.

Shannon, C. E. 1948.A mathematical theory of communication. *Bell System Technical Journal, vol. 27, pp. 379-423 y 623-656*.

Squire, D., W. Müller, H. Müller, T. Pun. 2000. *Content-based query of image databases: inspirations from text retrieval*. Pattern Recognition Letters. Selected Papers from The 11th Scandinavian Conference on Image Analysis SCIA '99, 21(13-14):1193-1198.

Yang, Y., J. O. Pedersen. 1997. *A Comparative Study on Feature Selection in Text Categorization*. Proceedings of ICML-97, 14th International Conference on Machine Learning.